

Pietsch, Marcus; Van den Ham, Ann-Katrin; Köller, Olaf

## **Wirkungen von Schulinspektion. Ein Rahmen zur theoriegeleiteten Analyse von Schulinspektionseffekten**

*Pietsch, Marcus [Hrsg.]; Scholand, Barbara [Hrsg.]; Schulte, Klaudia [Hrsg.]: Schulinspektion in Hamburg. Der erste Zyklus 2007 - 2013: Grundlagen, Befunde und Perspektiven. Münster [u.a.] : Waxmann 2015, S. 117-135. - (HANSE - Hamburger Schriften zur Qualität im Bildungswesen; 15)*



### **Quellenangabe/ Reference:**

Pietsch, Marcus; Van den Ham, Ann-Katrin; Köller, Olaf: Wirkungen von Schulinspektion. Ein Rahmen zur theoriegeleiteten Analyse von Schulinspektionseffekten - In: Pietsch, Marcus [Hrsg.]; Scholand, Barbara [Hrsg.]; Schulte, Klaudia [Hrsg.]: Schulinspektion in Hamburg. Der erste Zyklus 2007 - 2013: Grundlagen, Befunde und Perspektiven. Münster [u.a.] : Waxmann 2015, S. 117-135 - URN: urn:nbn:de:0111-pedocs-115319 - DOI: 10.25656/01:11531

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-115319>

<https://doi.org/10.25656/01:11531>

in Kooperation mit / in cooperation with:



**WAXMANN**  
[www.waxmann.com](http://www.waxmann.com)

<http://www.waxmann.com>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen. Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### **Terms of use**

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### **Kontakt / Contact:**

peDOCS  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Marcus Pietsch, Barbara Scholand,  
Klaudia Schulte (Hrsg.)

# Schulinspektion in Hamburg

Der erste Zyklus 2007–2013:  
Grundlagen, Befunde,  
Perspektiven



HANSE  
Hamburger Schriften zur Qualität  
im Bildungswesen

herausgegeben von  
Detlef Fickermann und Knut Schwippert

Band 15



Waxmann 2015  
Münster • New York

Marcus Pietsch, Barbara Scholand,  
Klaudia Schulte (Hrsg.)

# Schulinspektion in Hamburg

Der erste Zyklus 2007–2013:  
Grundlagen, Befunde und Perspektiven



Das gedruckte Buch finden Sie [hier](#).



Waxmann 2015  
Münster • New York

**Bibliografische Informationen der Deutschen Nationalbibliothek**

Die Deutsche Nationalbibliothek verzeichnet diese Publikation  
in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten  
sind im Internet über <http://dnb.d-nb.de> abrufbar.

**HANSE – Hamburger Schriften zur Qualität im Bildungswesen, Band 15**

ISSN 1864-2225

ISBN 978-3-8309-3278-9

© Waxmann Verlag GmbH, 2015  
Steinfurter Straße 555, 48159 Münster

[www.waxmann.com](http://www.waxmann.com)  
[info@waxmann.com](mailto:info@waxmann.com)

Umschlaggestaltung: Pleßmann Design, Ascheberg  
Umschlagfoto: © Robert Kneschke – Fotolia.de  
Lektorat und Satz: Judith Zimmer, Hamburg  
Druck: Mediaprint, Paderborn

Gedruckt auf alterungsbeständigem Papier,  
säurefrei gemäß ISO 9706

Alle Rechte vorbehalten. Nachdruck, auch auszugsweise, verboten.  
Kein Teil dieses Werkes darf ohne schriftliche Genehmigung des Verlages  
in irgendeiner Form reproduziert oder unter Verwendung elektronischer  
Systeme verarbeitet, vervielfältigt oder verbreitet werden.

# Inhalt

Vorwort. ....	9
---------------	---

## Teil 1: Die Grundlagen der Schulinspektion

*Norbert Maritzen*

Schulinspektion – Aspekte der Transformation von Governance-Strukturen . . .	13
--	----

*Norbert Maritzen & Herbert Altrichter*

Schulinspektion in Hamburg – ein Gespräch über Startbedingungen und Entwicklungen . . . . .	37
--	----

*Martina Diedrich*

Aufbau und Rolle der Schulinspektion Hamburg . . . . .	57
--	----

*Thomas Brüsemeister*

Erfahrungsbereiche der Hamburger Schulinspektion – Governance-analytische Programmtheorie . . . . .	77
--	----

*Susanne Ulrich*

Das Qualitätsleitbild der Schulinspektion Hamburg . . . . .	101
---	-----

## Teil 2: Die Arbeit der Schulinspektion

*Marcus Pietsch, Ann-Katrin van den Ham & Olaf Köller*

Wirkungen von Schulinspektion: Ein Rahmen zur theoriegeleiteten Analyse von Schulinspektionseffekten . . . . .	117
---	-----

*Moritz G. Sowada*

Expertenurteile – Achillesferse oder Trumpf der Schulinspektion? . . . . .	137
--	-----

*Knut Schwippert*

Daten für die Schulentwicklung – auf die Perspektive kommt es an . . . . .	157
--	-----

<i>Barbara Scholand unter Mitarbeit von Stephanie Graw</i>	
Wie Schulinspektor/innen fragen und Grundschüler/innen antworten . . . . .	177

<i>Peter Schulze &amp; Christiane von Schachtmeyer</i>	
Der Blick aus der Praxis – die Schulformexpertinnen und -experten . . . . .	219

### **Teil 3: Die Befunde der Schulinspektion**

<i>Melanie Ehren &amp; Jaap Scheerens</i>	
Evidenzbasierte Referenzrahmen zur Schulqualität als Grundlage von Schulinspektion . . . . .	233

<i>Holger Gärtner</i>	
Zusammenhang von Schul- und Unterrichtsqualität und schulischen Rahmenbedingungen . . . . .	273

<i>Marcus Pietsch &amp; Timo Ehmke</i>	
Die Qualität des Unterrichts an SINUS-Grundschulen. Eine Programmevaluation mithilfe von Daten der Schulinspektion Hamburg . . . . .	295

<i>Klaudia Schulte &amp; Markus Lücken</i>	
Der Einfluss schulischer Prozesse auf die Lernentwicklung der Schülerinnen und Schüler an weiterführenden Schulen in Hamburg . . . . .	317

<i>Susanne Peters</i>	
Eltern als Stakeholder von Schule. Erkenntnisse über die Sichtweise von Eltern durch die Hamburger Schulinspektion. . . . .	341

### **Teil 4: Die Zukunft der Schulinspektion**

<i>Sabine Müller</i>	
Zukunft der externen Evaluation – Trends und Perspektiven in Deutschland. .	369

<i>Hans-Jürgen Kuhn</i>	
Rechtliche, strukturelle und politische Rahmenbedingungen für externe Evaluation in Deutschland. . . . .	385

*Martina Diedrich*

Der zweite Zyklus der Schulinspektion Hamburg: ein Ausblick. . . . . 419

*Martin Spiewak, Angela Hellberg, Claudia Lemke, Gerrit Petrich,  
Norbert Rosenboom, Mina Zulal*

Die Perspektive der Stakeholder: Ein Interview von Martin Spiewak . . . . . 437

Autorinnen und Autoren . . . . . 455





## Vorwort

Vor rund zehn Jahren wurden die ersten Schulinspektorate in Deutschland eingeführt. Als Teil einer veränderten Steuerungsstrategie, die die Qualitätssicherung und -entwicklung von Schulen ins Zentrum setzt, sollen sie die Qualität schulischer Prozesse evaluieren, prozessuale Mindeststandards in Schule und Unterricht sichern und darüber hinaus Schulentwicklung stimulieren.

Die Hamburger Schulinspektion ist keine Inspektion der allerersten Stunde, wurde sie doch erst im Jahr 2007 offiziell eingerichtet. Zu diesem Zeitpunkt waren in vielen anderen Bundesländern schulische Evaluatorinnen und Evaluatoren bereits zwei bis drei Jahre im Feld und hatten die unterschiedlichsten Verfahren entwickelt und verschiedene Konzepte erprobt. Von ihren vielfältigen Erfahrungen mit diesem in Deutschland dahin unbekannten Instrument konnte die Hamburger Schulinspektion profitieren.

Gleichwohl hatte man in Hamburg bereits seit der Jahrtausendwende erste Konzepte zur externen Prozessevaluation an Einzelschulen in Hamburg diskutiert und entwickelt. Den Impuls hierfür gaben insbesondere die Ergebnisse der Längsschnittstudie LAU (Aspekte der Lernausgangslage und der Lernentwicklung), die seit Mitte der 1990er Jahre regelmäßig durchgeführt wurde. Es vergingen jedoch noch mehrere Jahre, bis die erste Hamburger Schule tatsächlich durch ein Inspektionssteam evaluiert wurde. Seitdem bewegt sich die Schulinspektion im Hamburger Schulsystem stets im Spannungsfeld von Administration, Praxis und Wissenschaft und muss dabei ihren eigenen sowie den Ansprüchen der verschiedenen schulischen und außerschulischen Stakeholder gerecht werden.

In ihrem ersten Zyklus, der im Jahr 2007 begann und bis zum Jahr 2013 andauerte, inspizierte die Schulinspektion 375 Hamburger Schulen. Während dieser sechs Jahre nahmen etwa 10.000 Lehrerinnen und Lehrer, gut 36.000 Eltern, rund 92.000 Schülerinnen und Schüler sowie etwa 2.000 Vertreterinnen und Vertreter Hamburger Unternehmen an den schriftlichen Befragungen der Inspektion teil. Darüber hinaus wurden fast 30.000 Unterrichtssequenzen durch die Inspektorinnen und Inspektoren beobachtet und etwa 2.000 Interviews mit Schulleitungen, Eltern- und Schülerräten sowie Lehrerkollegien und weiteren Mitarbeiterinnen und Mitarbeitern der inspizierten Schulen, aber auch mit Vertreterinnen und Vertretern von Ausbildungsbetrieben geführt.

Die Stärke des vorliegenden Bandes ist es, so meinen wir, dass er der besonderen Stellung der Schulinspektion im Gesamtgefüge des Hamburger Bildungssystems Rechnung trägt, indem er Akteure aus den verschiedenen Bereichen und mit unterschiedlichen Perspektiven zu Wort kommen lässt. Ein Teil der Beiträge ist empirisch ausgerichtet und analysiert die Arbeit der Inspektion oder nutzt die im Rahmen der Schulinspektion Hamburg erhobenen Daten für weiterführende

Analysen zum Hamburger Schulsystem. Andere Beiträge wiederum setzen sich mit den Rahmenbedingungen, Grundlagen und Folgen von Schulinspektion auseinander oder widmen sich den Erfahrungen und Erwartungen einzelner Akteure und Akteursgruppen.

Wir bedanken uns herzlich bei allen Autorinnen und Autoren dieses Bandes für die kooperative, kollegiale und anregende Zusammenarbeit. Insbesondere haben wir uns darüber gefreut, dass die vielen Kolleginnen und Kollegen aus diversen Kontexten und Arbeitsfeldern umstandslos bereit waren, ihre Expertise in Form fundierter und kritischer Beiträge sowie Gutachten beizusteuern. Sehr glücklich sind wir vor allem auch darüber, dass sich die Vertreterinnen und Vertreter der schulischen Stakeholder aktiv in diese Publikation eingebracht haben.

In der Zusammenarbeit mit all diesen verschiedenen Personen ist uns deutlich geworden, dass es institutionen- und akteursübergreifend ein großes Interesse daran gibt, über gute Bildung in und für Hamburg ins Gespräch zu kommen und sie gemeinsam zu gestalten. Dafür bieten – nicht nur, aber auch – die Schulinspektion und ihre Ergebnisse konstruktive Gesprächsanlässe. In diesem Sinne wünschen wir Ihnen eine ertragreiche Lektüre!

Hamburg im Juni 2015

Marcus Pietsch, Barbara Scholand & Klaudia Schulte

# Wirkungen von Schulinspektion: Ein Rahmen zur theoriegeleiteten Analyse von Schulinspektionseffekten

*Marcus Pietsch, Ann-Katrin van den Ham & Olaf Köller*

## *Zusammenfassung*

*Schulinspektionen sollen die Qualität von Schule und Unterricht verbessern und hierdurch zur Steigerung von Schülerleistungen beitragen. Um die derzeitig gängigen Schulinspektionsverfahren wissensbasiert weiterzuentwickeln, intendierte Wirkungen mit höherer Wahrscheinlichkeit als bislang zu erzielen und somit die Wirksamkeit der Verfahren zu erhöhen, ist es notwendig, empirische Evidenz zu wirksamkeitsförderlichen Bedingungen und Vorgehensweisen zu sammeln. Eine gute Möglichkeit, dies zu erforschen, bieten theoriebasierte Evaluationen. Da die bisherige Forschung zu Wirkungen von Schulinspektionen eher kursorisch erfolgt, wird im Rahmen des vorliegenden Beitrages ein dreischrittiges Verfahren zur systematischen Evaluation von Schulinspektionseffekten vorgeschlagen. Hierbei wird insbesondere dafür plädiert, dass im Rahmen theoriegeleiteter Evaluation von Schulinspektionseffekten die Komplexität und die Kompliziertheit der Intervention stärker berücksichtigt werden als bisher und die im Rahmen programmtheoretisch fundierter Analysen generierten Befunde mittels des Ansatzes des Interpretation/Use Argument nach Kane validiert werden, um so die Belastbarkeit der Studien und der daraus abzuleitenden Konsequenzen für die Weiterentwicklung von Inspektionsverfahren zu erhöhen.*

## 1. Einleitung

Schulinspektionen sollen dazu beitragen, Schülerleistungen zu verbessern. Dies, so die Erwartung, soll gelingen, indem sie Schulen durch die Rückmeldung beobachteter Stärken und Schwächen Impulse für die Weiterentwicklung von Schule und Unterricht liefern. Hierfür wird nach normativen Vorgaben, die in der Regel in landesspezifischen Qualitätsrahmen oder Qualitätstableaus formuliert wurden, durch Schulinspektoren extern an Schulen evaluiert. Es wird dann davon ausgegangen, dass Schulverantwortliche die zurückgemeldeten Informationen für eine wissensbasierte Schul- und Unterrichtsentwicklung nutzen, die sich wiederum in verbesserten Schülerleistungen niederschlägt. Wie Pietsch, Schulze, Schnack und Krause (2011) herausstellen, knüpfen die diesbezüglichen Wirksamkeitserwartungen an Schulinspektionen vor allem an die Forschung zum zielorientierten

Feedback (vgl. Kluger & DeNisi 1996; Ramaprasad 1983; Visscher & Coe 2003) an. Entsprechend wird erwartet, dass das empirisch begründete Aufzeigen von Differenzen zwischen normativ vorgegebenen Soll- und empirisch beobachteten Ist-Zuständen dazu führt, dass in extern evaluierten Schulen infolge der Rückmeldung eine Handlungsoptimierung geplant wird, die es ermöglicht, anzustrebende Ziele in Zukunft besser zu erreichen.

Eine Vielzahl von Studien zur Wirksamkeit von Schulinspektionen zeigt jedoch, dass dieses Ziel häufig nicht erreicht wird oder sogar nicht-intendierte Nebeneffekte beobachtbar sind. Die Befundlage ist hierbei sehr heterogen. So zeigen einzelne Studien, dass sich Effekte auf die innerschulische Qualitätsentwicklung finden lassen, und andere Studien negieren genau diesen Befund (vgl. zur Übersicht z. B. de Wolf & Janssens 2007; Dederig 2012; Husfeldt 2011). Auch mit Blick auf die Verbesserung von Schülerleistungen weisen bislang nur zwei Studien aus den Niederlanden und Deutschland positive Effekte nach (vgl. Pietsch et al. 2014; Luginbuhl et al. 2009). Diejenigen Studien hingegen, die sich auf Effekte der englischen Schulinspektion OFSTED (*The Office for Standards in Education*) beziehen, zeigen, dass bezogen auf das Gesamtsystem keine oder sogar leicht negative Wirkungen auf Schülerleistungen als Folge von Schulinspektionen beobachtet werden können (vgl. de Wolf & Janssens 2007). Darüber hinaus wird in einzelnen Studien von nicht-intendierten Nebenwirkungen bei Schulinspektionen berichtet. Demnach steigt den Befunden zufolge das Stresserleben bei Lehrkräften während einer Inspektion (vgl. z. B. Brunson et al. 2006) oder es findet eine überzogene positive Selbstdarstellung der Schule (*Window Dressing*) statt (vgl. de Wolf & Janssens 2007).

Insgesamt findet sich somit auch zehn Jahre nach Einführung der ersten Schulinspektorate in Deutschland eine uneindeutige Befundlage, die keine kohärente Einschätzung darüber ermöglicht, ob Schulinspektionen ein Instrument sind, von dem erwartet werden kann, dass es zu einer Verbesserung von Schul- und Unterrichtsqualität und infolge dessen zu einer Steigerung von Schülerleistungen führt. Pietsch et al. (2014) haben vorgeschlagen, diesem Problem mithilfe einer Kombination kausalanalytischer Blackbox-Verfahren und theoriegeleiteter Evaluationen zu begegnen. Während die erste Herangehensweise Aufschluss darüber geben kann, ob Inspektion grundsätzlich die intendierten Effekte erzielt oder nicht, ist letztere insbesondere dann sinnvoll, wenn es um Entscheidungen zur weiteren Gestaltung, zu Optimierung, Veränderung oder Verbesserung von Inspektionsverfahren geht. Einen Punkt, den bereits Cronbach et al. (1980, S. 251) als besonders relevantes Kernelement von Evaluationen herausstellen:

„Knowing this week’s score does not tell the coach how to prepare next week’s game. The information that an intervention had satisfactory or unsatisfactory outcomes is of little use by itself; users [...] need to know what led to success or failure.“

Während die Evaluation mithilfe von Blackbox-Verfahren zwar methodisch anspruchsvoll, ansonsten jedoch geradlinig und überschaubar ist, erfordert eine theoriegeleitete Evaluation von Inspektionswirkungen eine komplexere Herangehensweise, um zu belastbaren Resultaten zu gelangen (vgl. Astbury & Leeuw 2010). Im Folgenden wird daher ein Rahmen für eine theoriegeleitete Analyse von Schulinspektionseffekten vorgestellt, die es ermöglichen soll, Befunde zur Wirkung von Schulinspektionen systematisch zu generieren. Hierzu werden zuerst Annahmen zur Wirksamkeit und den Funktionen von Schulinspektionen beschrieben. Anschließend wird dargestellt, welche Gründe es aus programmtheoretischer Perspektive dafür geben kann, dass Schulinspektionen ggf. nicht wirksam werden, um anschließend auf ein strukturiertes Vorgehen zur Evaluation von Schulinspektionswirkungen einzugehen.

## 2. Annahmen zur Wirksamkeit und zu Funktionen bei Schulinspektionen

### 2.1 Funktionen von Schulinspektionen

Schulinspektion hat die Aufgabe, Informationen aus dem Bereich Bildung und Erziehung systematisch zu beschaffen und aufzubereiten, um auf diesem Wege Akkreditierungen, Rechenschaftslegungen und Diagnosen für systemisches Lernen im Bildungssystem zu ermöglichen. Schulinspektionen sollen entsprechend dazu beitragen, elementare Standards von Bildungsqualität zu gewährleisten, einen verbesserten Service für das einzelschulische Qualitätsmanagement zu bieten und unterschiedlichen Akteuren im Bildungssystem verwertbare Informationen zu schulischen Prozessqualitäten bereitzustellen (vgl. Döbert et al. 2008; Maritzen in diesem Band; Pietsch et al. 2009, 2013). Dabei liegt der Schwerpunkt der Inspektionsarbeit auf der Evaluation einzelschulischer Prozesse (vgl. Böttcher & Kotthoff 2010; Döbert et al. 2008; Maritzen in diesem Band; van Ackern & Klemm 2009), wobei die Inspektionen in den Ländern trotz augenscheinlicher Unterschiede in Ausstattung, Ausgestaltung und ministerieller Anbindung „auf Grundlage von Verfahren mit wiederkehrenden Standardelementen“ erfolgen (Maritzen in diesem Band, S. 16).

Grundlage für die externe Einzelschulevaluation bilden dabei in allen Ländern die landesspezifischen Qualitätsrahmen oder Qualitätstableaus, die im Sinne eines Input-Prozess-Output-Modells (teilweise unter Berücksichtigung von Kontextfaktoren) Anforderungen an Schulqualität definieren (vgl. Ehren & Scheerens in diesem Band; Müller 2010). In allen Inspektionen wurden, basierend auf diesen Qualitätskatalogen, Leistungsindikatoren definiert, die es ermöglichen sollen, die Qualität einzelner Prozesse möglichst differenziert zu erfassen (vgl. Böttcher

& Kotthoff 2007). Geschulte Experten – meist Lehrkräfte, Schulleitungen oder Schulaufsichten – evaluieren anhand dieser Indikatoren die Schulen in Teams von zwei bis vier Personen (vgl. Döbert et al. 2008; Müller 2010). Die Datenerhebung selbst erfolgt über ein umfangreiches Repertoire an Methoden. Fragebögen und Interviews, gerichtet an verschiedene Schulbeteiligte, sowie Schulbegehungen und die Begutachtung schulinterner Dokumente und Statistiken gehören derzeit ebenso wie die Beobachtung von Unterrichtseinheiten zum länderübergreifenden Methodenstandard (vgl. Döbert et al. 2008). Die derart ermittelten Befunde werden in einem Bericht zusammengefasst und den Schulen sowie ggf. weiteren Beteiligten, wie z. B. Schulaufsichten, übergeben. In regelmäßigem Turnus werden darüber hinaus Berichte verfasst, die Befunde der Schulinspektionen auf Systemebene präsentieren.

Durch diese Rückmeldungen aus Schulinspektionen erhalten die Schulbeteiligten, ebenso wie Bildungsadministration und Bildungspolitik, Informationen zur extern wahrgenommenen Qualität innerschulischer Prozesse auf verschiedenen Ebenen. Obwohl dies häufig nicht explizit benannt wird, steht hinter dem Vorgehen, aufbereitete Informationen an verschiedene Gruppen auf Schul- und Schulsystemebene zurückzumelden, die Idee, dass die Rückmeldungen von Evaluationsergebnissen als Grundlage für Entwicklungen an den evaluierten Schulen genutzt werden, die in ihrer Konsequenz in verbesserte fachliche Schülerleistungen münden (vgl. de Wolf & Janssens 2007; Ehren & Visscher 2006; Ehren et al. 2005). Insbesondere in den deutschen Ländern zeichnet sich diesbezüglich aktuell ein Trend ab (vgl. Böttcher & Kotthoff 2010, S. 307), „demzufolge Schulinspektionen [...] als ein Unterstützungs- und Dienstleistungsangebot für die Schule verstanden werden, deren Erkenntnisse wirksam für die interne Schulentwicklung genutzt werden können bzw. sollen“. Folgerichtig betont Köller (2009), dass der Wert von Schulinspektionen insbesondere darin läge, dass sie Hinweise auf die Prozesse vor Ort gäben, mithilfe derer Hypothesen darüber generiert werden könnten, welche Veränderungen von Schule und Unterricht zu höheren Lernerträgen der Schülerinnen und Schüler führen können.

## 2.2 Annahmen zur Nutzung von Inspektion als Grundlage für Schulentwicklung

Entsprechend wird davon ausgegangen, dass Schulverantwortliche die Inspektion und ihre Berichte nutzen, um hieraus konkrete Maßnahmen für die Schul- und Unterrichtsentwicklung abzuleiten (vgl. Ehren & Visscher 2006, 2008). Im Zentrum eines solch entwicklungsfokussierten, anwenderorientierten Verständnisses von Evaluation, wie sie den Schulinspektionsverfahren in den deutschen Bundesländern zugrunde liegt, steht daher zwangsläufig der Evaluationsnutzen, der sich aus

der Untersuchung der Qualität von Schule und Unterricht für die Schulbeteiligten ergibt. Dabei ist es von grundlegender Bedeutung zu wissen, in welcher Art und Weise Informationen aus Schulinspektionsverfahren innerhalb von Schulen genutzt werden können, wenn das Ziel Qualitäts- und Schulentwicklung heißt, da „das Verwendungsinteresse der [...] Anwendersysteme [...] unterschiedlich ist, je nachdem ob die Evaluation eher zur Überprüfung/Kontrolle, zur Legitimation, zur Innovation/Entwicklung oder lediglich zur Bestätigung durchgeführt wird“ (vgl. Stamm 2003, S. 195 f.). Diesbezüglich lassen sich mit Blick auf die Nutzung von Evaluationsergebnissen grundsätzlich drei Nutzungsformen voneinander unterscheiden, die je einen spezifischen Zweck verfolgen (vgl. King & Pechman 1984; Weiss 1998):

- 1) Instrumentelle Nutzung (*Instrumental Use*): Die instrumentelle Nutzung von Informationen aus Evaluationen bezieht sich darauf, konkrete Probleme zu lösen bzw. konkrete Entscheidung zu treffen, die den Evaluationsgegenstand betreffen.
- 2) Konzeptionelle Nutzung (*Conceptual Use*): Die konzeptionelle Nutzung von Informationen aus Evaluationen findet statt, wenn diese indirekt genutzt werden, um das Wissen bzgl. des Evaluationsgegenstandes zu erweitern.
- 3) Symbolische Nutzung (*Symbolic Use*): Die symbolische Nutzung von Informationen aus Evaluationen findet statt, wenn Befunde zum Evaluationsgegenstand eingesetzt werden, um bereits getroffene Entscheidungen gegenüber Dritten zu legitimieren.

Dabei dient die instrumentelle Nutzung primär dem Zweck der konkreten und zielgerichteten Weiterentwicklung, die konzeptionelle Nutzung wiederum hat die indirekte Veränderung von Einstellungen, Meinungen und Kognitionen der an der Evaluation Beteiligten zum Ziel und die symbolische Nutzung dient der Bekräftigung von Positionen in Diskursen und argumentativen Auseinandersetzungen (vgl. Weiss 1998). Für die Nutzer der Evaluationsbefunde auf Ebene der Einzelschule, müssen die evaluativ gewonnen Informationen daher, wenn es um die konkrete Weiterentwicklung von Schule und Unterricht geht, in erster Linie einen instrumentellen, einen direkten praxisbezogenen Nutzen zur Entscheidungsfindung erbringen (vgl. Scheerens 2007). Einen Nutzen also, der es ermöglicht, inkrementelle Entscheidungen zu treffen, um so z.B. ein bestimmtes Programm zu beenden, es zu modifizieren, Aktivitäten zu verändern oder Weiterbildungsmaßnahmen zu gestalten (vgl. Weiss 1998). Eine solche instrumentelle Nutzung von Evaluationsbefunden kann jedoch nur unter einer der vier folgenden Voraussetzungen stattfinden (vgl. Weiss 1980): 1) wenn die Implikationen, die sich aus den Befunden ergeben, nicht kontrovers sind; 2) wenn die geplanten Veränderungen sich innerhalb des bisher geplanten Programms bewegen und nicht allzu umfassend sind; 3) wenn das

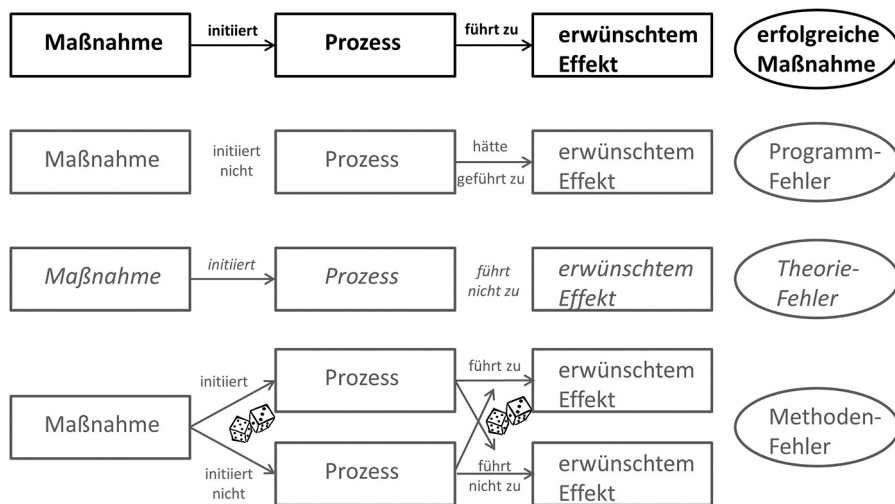


Umfeld des Programms stabil ist, es keine größeren Veränderungen im Bereich von Leitung, Finanzen und Teilnehmern gibt oder wenn 4) ein Programm in der Krise ist und niemand weiß, was getan werden muss. Die direkte instrumentelle Nutzung, die in Veränderungen und Entwicklung mündet, findet somit vor allem dort statt, wo der bestehende Status quo nicht gestört wird (vgl. Nutley et al. 2003).

### 3. Problemstellung

Wirksamkeits- und Wirkungsstudien untersuchen nun, ob Schulinspektionen durch Schulverantwortliche genutzt werden und welche Wirkungen erreicht werden. Intendiert ist eine instrumentelle Nutzung, die zu einer Verbesserung von Schülerleistungen und diesbezüglich vermittelnden Faktoren der Schul- und Unterrichtsqualität führt. Lassen sich diese Zusammenhänge nicht nachweisen, so kann dies theoretisch auf drei Gründe zurückgeführt werden (vgl. Pietsch et al. 2013; Stame 2010): Erstens kann ein Programm- resp. Implementations-, zweitens ein Theorie- und/oder drittens ein Methodenfehler vorliegen (vgl. Abb. 1).

Abbildung 1: Wirkungskette und mögliche Ursachen für Unwirksamkeit aus programm-theoretischer Perspektive



Während ein Programmfehler vorhanden ist, wenn es mittels einer Intervention nicht gelingt, eine intendierte Wirkung nachweisbar zu erzeugen, liegt ein Theoriefehler vor, sofern ein theoretisch postulierter Wirkungszusammenhang nicht ausreichend valide begründet wurde. Darüber hinaus kann eine nicht nachweisbare

Wirkung letztlich auch darauf zurückzuführen sein, dass Methodenfehler vorliegen. Die Nicht-Wirksamkeit einer Intervention rührt in diesem Falle daher, dass es im Rahmen der empirischen Kausalstudie aufgrund methodologischer Unzulänglichkeiten nicht gelingt, den angenommenen Wirkungszusammenhang empirisch verlässlich zu überprüfen.

Empirische Untersuchungen lassen sich zu allen drei Bereichen finden, wobei derzeit ein grundsätzliches Problem darin besteht, dass Programmtheorien, die eine theoriebasierte Evaluation von Inspektionswirksamkeit ermöglichen sollen, zu unterkomplex sind und die Kompliziertheit der Intervention Schulinspektion nicht berücksichtigen (vgl. de Wolf & Janssens 2007; Pietsch et al. 2013). Dies wiederum hat zur Folge, dass Aussagen zu Programmfehlern – also dazu, ob Schulinspektionen ggf. nicht zufriedenstellend implementiert sind und/oder ihrer Aufgabe nur unzureichend nachkommen – nur schwer möglich sind und die bislang vorliegenden Befunde daher nur wenig verlässliche Aussagen zur Wirksamkeit von Schulinspektionen auf die Entwicklung von Schule und Unterricht sowie auf Schülerleistungen ermöglichen. Entscheidungen zum weiteren Umgang mit Inspektionsverfahren sind daher aktuell nur schwer möglich.

Um die Wirkungen von Inspektionen systematisch zu erforschen, schlagen wir daher ein dreischrittiges Verfahren vor, das der Komplexität des Untersuchungsgegenstandes gerecht zu werden versucht. Hierbei soll in einem ersten Schritt gewährleistet werden, dass das Zielkriterium, der Effekt, um den es geht, sowie der logisch angenommene Wirkungszusammenhang vor Beginn einer Untersuchung klar und eindeutig definiert wird. Im zweiten Schritt geht es darum, die Wirkmechanismen im Rahmen eines logischen Modells zu beschreiben und zu untersuchen, das der Komplexität und der Kompliziertheit des Untersuchungsgegenstandes gerecht wird. In einem dritten Schritt sollen die Ergebnisse dieser Untersuchungen zusammengeführt werden, um hieraus belastbare Konsequenzen auf Systemebene, aber auch auf Ebene der einzelnen Inspektion ziehen zu können.

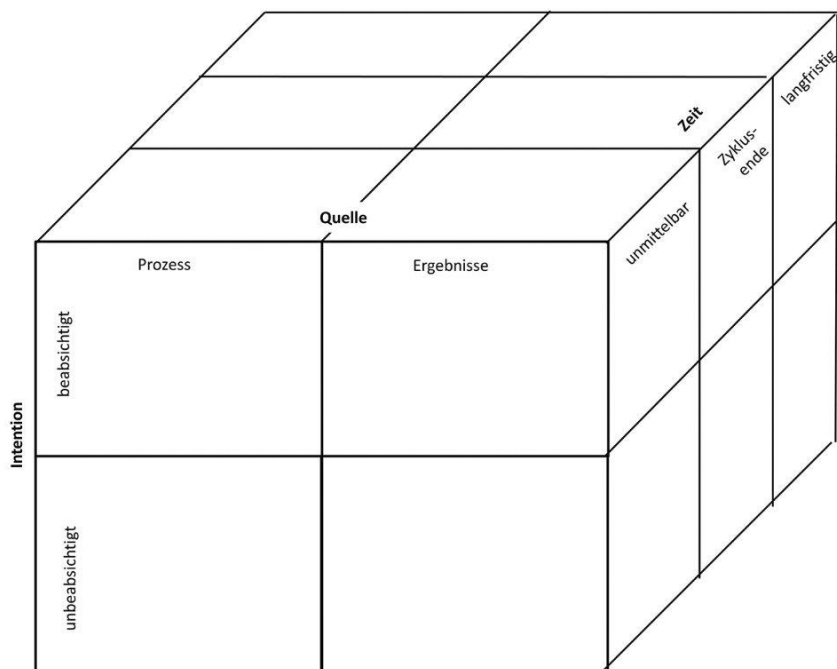
#### 4. Ein konzeptioneller Analyserahmen zur Evaluation von Schulinspektionswirkung

##### 4.1 Grundlage: Definition von Inspektionswirkungen mithilfe eines mehrdimensionalen Modells

Schulinspektionen können, wie beschrieben, sowohl intendierte als auch nicht-intendierte Wirkungen nach sich ziehen, aber auch vollkommen wirkungslos bleiben. Dabei geht es um den Einfluss (im fachwissenschaftlichen Diskurs wird diesbezüglich von ‚*Influence*‘ gesprochen, vgl. z. B. Herbert 2014), den Inspektio-

nen auf Schule, Unterricht und Schülerleistungen haben. Daher ist es im Rahmen einer theoriegeleiteten Evaluation von Schulinspektionswirkungen in einem ersten Schritt notwendig, zu klären, welche Wirkung in welchem Zeitraum generiert und durch welchen Prozess (Durchführung einer Inspektion oder Rückmeldung von Ergebnissen) erwartet wird. Möchte man also den Einfluss von Schulinspektionen auf Schülerleistungen sowie vermittelnde Faktoren wie Schul- und Unterrichtsentwicklung analysieren, benötigt man entsprechend elaborierte Modelle. Eben solche haben Kirkhart (2000) sowie Alkin und Taut (2003) vorgeschlagen. Mithilfe dieser dreidimensionalen Modelle lässt sich die Wirkung von Evaluationen – und entsprechend auch von Schulinspektionen – fassen (vgl. Abb. 2).

Abbildung 2: Modell zur Definition von Inspektionswirkungen (nach Kirkhart 2000)



Der Vorteil eines solchen Schemas liegt vor allem darin, dass er über den Begriff der Nutzung hinaus geht und es vor allem auch erlaubt, nicht-intendierte Wirkungen in die Analysen mit einzubeziehen und darüber hinaus Effekte resp. Veränderungen – und somit Wirkungen, nicht die Nutzung von Evaluationen – in den Mittelpunkt stellt. Weiterhin ermöglicht dieses Schema, zu berücksichtigen, dass Schulinspektionen, anders als z. B. Vergleichsarbeiten und Lernstandserhebungen, auch einen Einfluss durch das Wirken von Inspektorinnen und Inspektoren vor Ort

haben und nicht ausschließlich Effekte durch die Rückmeldung von Ergebnissen erzielen. Die Stärken eines solchen Modells stellt Herbert (2014, S. 394) entsprechend heraus:

- „influence provides a definition and a framework that reflects the full impact of evaluation and a cohesive way to organize theoretical and empirical knowledge of the effect evaluation can have on programs;
- by adopting this more comprehensive view, influence allows for the study of implicit mechanisms that affect change, including processes at the individual, interpersonal, and collective levels;
- influence frameworks are oriented around linkages to more developed constructs in other fields of literature such as attitude change, priming, skill acquisition, and persuasion;
- shifting to an influence framework allows for the study of pathways of influence and the study of situations where evaluation failed to affect change; and influence is built around social betterment as the ultimate goal of evaluation, rather than use.“

Untersucht werden kann in einem solchen Design somit: a) ob intendierte Effekte durch die Evaluation erzielt wurden oder nicht (*Intention*), b) ob diese durch die Ergebnissrückmeldung oder den Evaluationsprozess zustande gekommen sind (*Quelle*) und c) ob die Effekte unmittelbar, zum Ende eines Zyklus oder aber erst langfristig nachweisbar sind (*Zeit*).

#### 4.2 Analyse: Theoriebasierte Evaluation mithilfe komplexer logischer Modelle

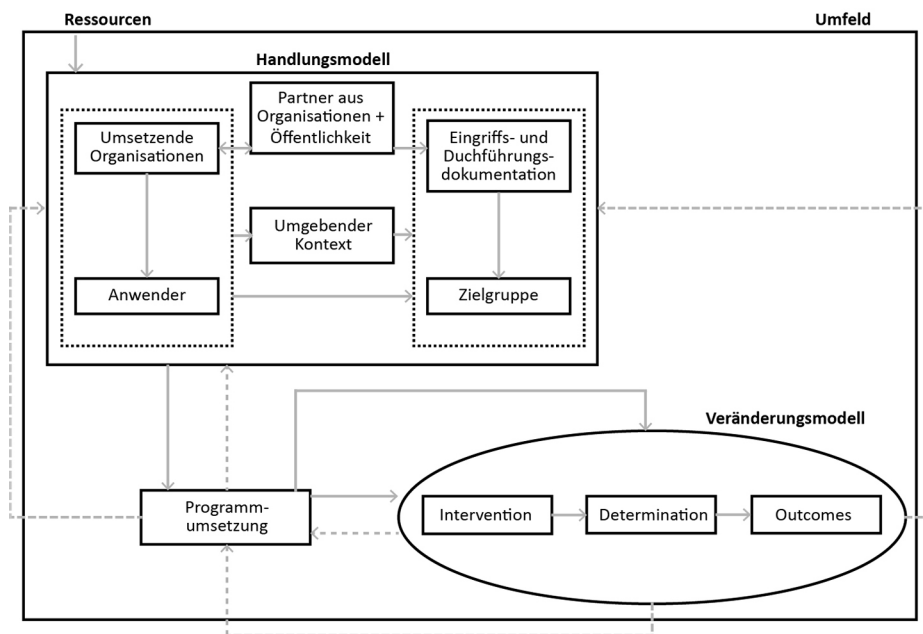
In einem zweiten Schritt geht es im Rahmen einer theoriegeleiteten Evaluation zu Wirkungen von Schulinspektion dann darum zu beschreiben, mithilfe welcher Mechanismen, auf welche Art und Weise die definierten Wirkungen durch ein Schulinspektionsverfahren generiert werden sollen. Von einfach-linearen Modellen wird abgeraten, da diese für die Analyse von Schulinspektionswirkungen zu unterkomplex sind: „The complexity and situatedness of much organizational activity begs for a style of theory that preserves some ambiguity.“ (vgl. Weber 2006, S. 120)

Komplexe programmtheoretische Modelle bestehen in der Regel (vgl. Chen 2006; Coryn et al. 2011) aus mindestens zwei Teilen (vgl. Abb. 3), die eng miteinander verknüpft sind und es u. a. ermöglichen, Kontextbedingungen und/oder rekursive Wirksamkeitsmechanismen zu inspizieren. Chen (2006) unterscheidet diesbezüglich zwischen einem Veränderungs- und einem Handlungsmodell:

„The action model and change model are closely related to each other and are essential for the success of a program. On the one hand, a change model is needed to justify the selection of an intervention for achieving the goals and/or outcomes and it provides a basis for developing the action model. On the other hand, the action model provides a blueprint to organize program activities and to activate and energize the change model for achieving program goals.“

Dabei umfasst das *Veränderungsmodell* drei Komponenten: a) die Intervention, die sich auf das Programm bezieht, mit dessen Hilfe Veränderungen bewirkt werden sollen, b) die Determination, die diejenigen Prozesse und Mechanismen umfasst, die den Input durch die Intervention, mit Blick auf mögliche Effekte, vermitteln, und c) die Outcomes, womit die antizipierten Effekte des Programms gemeint sind. Das Veränderungsmodell unterstellt entsprechend, dass die Implementation einer Intervention diejenigen Determinanten kausal beeinflusst, die ihrerseits zu Veränderungen in den Outputs führt.

Abbildung 3: Modell zur Evaluation von Inspektionswirkungen (nach Chen 2006)



Das *Handlungsmodell* hingegen beschreibt die Voraussetzungen und Annahmen der Intervention selber, stellt also dar, mit welchen Mitteln, unter welchen Voraus-

setzungen und bei welchen Adressaten die Intervention Wirkung erzielen soll. Das Handlungsmodell umfasst dabei sechs Bereiche:

- 1) *Umsetzende Organisationen*: Die Organisationen, die für die Koordination von Personal und die Ressourcenallokation bzgl. der Intervention verantwortlich sind.
- 2) *Anwender*: Die Personen, die für die konkrete Umsetzung resp. Implementierung des Programms verantwortlich sind.
- 3) *Partner aus Organisationen und Öffentlichkeit*: Die Organisationen oder Partner, die zusätzlich zur Primärorganisation notwendig sind, um die Intervention zielgerichtet durchzuführen.
- 4) *Umgebender Kontext*: Der ökologische Kontext, der direkt mit dem zu implementierenden Programm interagiert, z. B. in Form von Zielvorgaben, Normen, Gesetzen.
- 5) *Eingriffs- und Durchführungsdokumentation*: Der Umsetzungsplan, der die Intervention inkl. ihrer spezifischen Inhalte und Aktivitäten beschreibt und dabei die Schrittfolge definiert, mit der die Intervention im Feld umgesetzt werden soll.
- 6) *Zielgruppe*: Die Zielgruppe, für die das intervenierende Programm gedacht ist, insbesondere das Vorhandensein von klar definierten Ansprechpartnern und die Einbindung in das Programm.

#### 4.3 Konsequenzen: Validierung mithilfe von *Interpretation/Use-Argumentationen*

Die Ergebnisse, die mit diesen Modellen generiert werden können, müssen in einem letzten Schritt sinnvoll zusammengeführt und validiert werden, um hieraus Konsequenzen auf Systemebene, aber auch auf Ebene der einzelnen Inspektion zu ziehen. Scriven (2012, S. 2 ff.) betont konsequenterweise, dass Validität im Rahmen von Evaluation das zentrale Kriterium ist:

„Validity: This is the key criterion – the matter of truth. [...] It requires some evidence of ‚real‘ value, which usually (not quite always) means visible or directly testable evidence somewhere along the line of implications of the evaluation.“

Hierbei geht es dann vor allem darum, die Annahmen zur Wirkungsweise von Inspektionen zu prüfen. Was aber meint Validierung hier konkret? In den Standards für Pädagogisches und Psychologisches Testen von 1999 und 2014, die fachlich-methodische und ethische Richtlinien unter anderem für Validität beinhalten, wird Validität als das Ausmaß definiert, in dem empirische Nachweise und Theorien

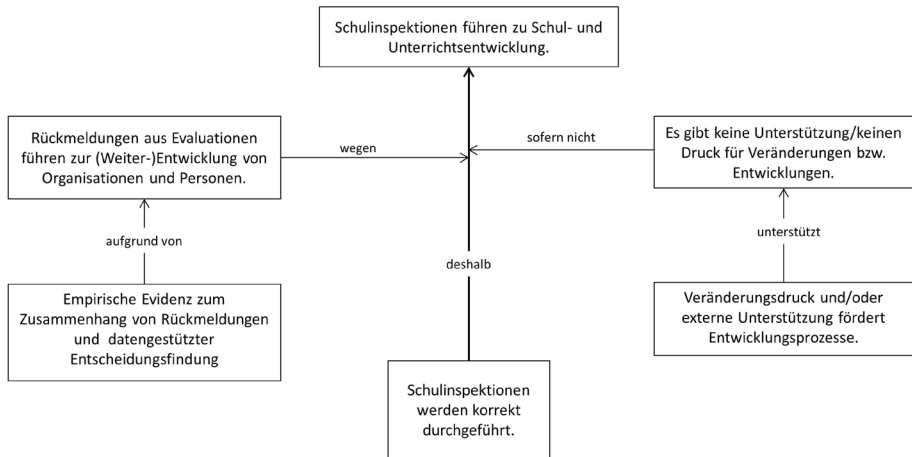
die Interpretation von Testergebnissen für eine bestimmte Nutzung stützen. Die Standards unterscheiden fünf Kategorien der Evidenz (Testinhalt, Antwortprozesse, Interne Struktur, Zusammenhänge mit anderen Variablen und Testkonsequenzen), die bei der Validierung evaluiert werden können. Jedoch wird Validität als ein einheitliches Konzept definiert (American Educational Research Association, American Psychological Association, National Council on Measurement in Education, 1999; S. 14):

„It is the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use.“

Zu validieren bedeutete in diesem Sinne, Argumente für und gegen die geplante Interpretation von Testergebnissen oder, im Falle von empirischen Untersuchungen zur Wirksamkeit von Schulinspektionen, von Studienergebnissen, zu konstruieren und zu evaluieren. Damit spiegeln die Standards den professionellen Konsens darüber wider, dass Validieren das Sammeln von Evidenz für eine wissenschaftlich solide Basis der Testwertinterpretationen und Schlussfolgerungen beinhaltet. In diesem Sinne ist Validität graduell. Für unterschiedliche Tests werden unterschiedliche Arten der Evidenz benötigt, wobei Validität ein einheitliches Konstrukt bleibt. Dabei ist Validität keine Eigenschaft des Tests oder der Untersuchung, sondern bezieht sich auf deren Interpretationen (vgl. Kane 2006; Shaw & Crisp 2012; Messick 1989; American Educational Research Association, American Psychological Association, National Council on Measurement in Education 1999 und 2014).

Neuere Validierungsansätze, die Validitätsnachweise in Form von Validitätsargumenten organisieren, wie der *Argument Based Approach* von Kane (2013), das *Evidence-Centered Design* von Mislevy (2003) und das *Assessment Use Argument* von Bachman (2005), basieren auf Toulmins Beschreibung von „Informellen Argumenten“ (1958, 2003). Die informellen Argumente werden in Form einer Argumentationskette aufgebaut, die für eine bestimmte Schlussfolgerung plädiert. In dieser Argumentationskette werden die Plausibilität von Behauptungen und Beobachtungen begründet und Zusammenhänge zwischen diesen bewiesen. In seinem Schema schließt der Argumentierende aufgrund von Daten und mithilfe einer Schlussregel auf die Konklusion. Dabei liefert eine Stützung Beweise für die Schlussfolgerung, Qualifikatoren bringen die Stärke der Behauptung zum Ausdruck und eine Ausnahme beschränkt oder verstärkt die Behauptung.

Abbildung 4: Modell zur argumentativen Validierung von Schulinspektionswirkungen  
(nach Toulmin 2003)

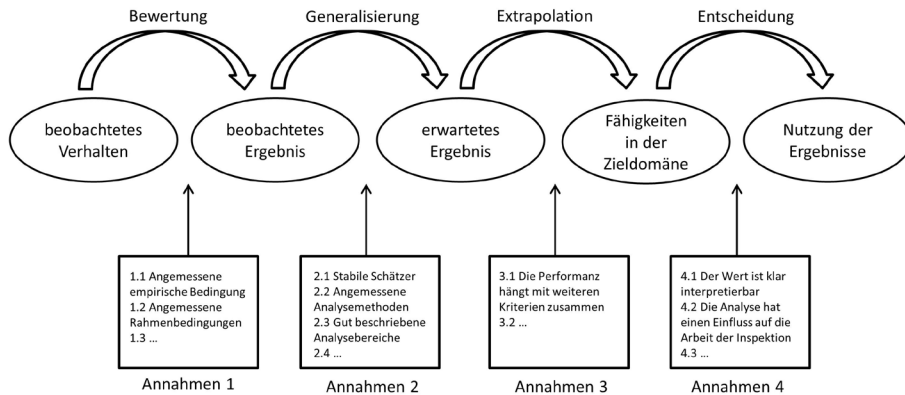


Ein Beispiel für so eine Behauptung wäre, dass Schulinspektion zu einer Verbesserung der Schul- und Unterrichtsqualität führt (siehe Abb. 4). Das Argument hierfür lautet, dass sie korrekt durchgeführt werden. Die Schlussregel ist eine Generalisierung, die genutzt wird, um die spezifischen Daten mit der spezifischen Behauptung zu verbinden. Im Falle des Beispiels in Abbildung 4 lautet die Schlussregel, dass Schulinspektionen, die korrekt durchgeführt werden, in der Regel zu einer Verbesserung der Schul- und Unterrichtsqualität an evaluierten Schulen führen. Die Stützung liefert Beweise für die Schlussfolgerung in Form von Theorien, Untersuchungen, Erfahrung etc.; alternative Erklärungen relativieren eine Behauptung und werden durch Ausnahmen beschränkt oder verstärkt.

Kane geht noch weiter als Toulmin (2003) bzw. Mislevy (2003) und argumentiert, dass für die Verbindung der Beobachtung mit der angestrebten Behauptung mehrere Arten von Schlussfolgerungen notwendig sind. In Abbildung 5 wird das sogenannte *Interpretation/Use Argument* (IUA) dargestellt, welches die multiplen Schlussfolgerungen und Ergebnismutzungen spezifiziert. Jede dieser Schlussfolgerungen basiert auf Annahmen, die Stützung benötigen. Die erste Schlussfolgerung „Bewertung“ geschieht bei der Ergebnisermittlung. Das beobachtete Verhalten bei oder infolge der Durchführung einer Inspektion wird in ein beobachtetes Ergebnis umgesetzt.



Abbildung 5: Interpretation/Use Argument zur Validierung von Schulinspektionswirkungen (nach Kane 2002)



Dabei wird beispielsweise angenommen, dass die Bewertungsprozeduren angemessen sind, korrekt angewendet wurden und frei von offenkundigen Fehlern sind. Die zweite Schlussfolgerung bezieht sich auf die Generalisierung der Ergebnisse aus einer ausgewählten Messung/Studie (mit bestimmten Kriterien, an einem bestimmten Tag, zu einer bestimmten Zeit) auf andere, vergleichbare Situationen. Dabei wird unter anderem davon ausgegangen, dass die Stichprobe der Umstände, unter denen gemessen wird, repräsentativ ist. Oft werden die Interpretationen auf weitere Bereiche wie z.B. die Realsituation erweitert. Die hierauf gründenden Schlussfolgerungen ermöglichen eine deskriptive Interpretation der Ergebnisse. Die Einbeziehung von Verwendungen und Entscheidungen führt schließlich zu einer entscheidungsbasierten Interpretation (Kane 2002). Bei der Schlussfolgerung „Entscheidung“ werden die Ergebnisse für das Treffen von Konsequenzen verwendet (Kane 2013).

## 5. Fazit

Vor rund zehn Jahren wurden die ersten Schulinspektorate in Deutschland eingeführt. Die Frage, welche Wirkungen Schulinspektionen nach sich ziehen, ist bislang jedoch nur in Ansätzen und darüber hinaus sehr unsystematisch erforscht. Befunde zur Wirksamkeit, also bezogen darauf, ob Schulinspektionen intendierte Wirkungen erzielen, liegen noch seltener vor. Problematisch ist darüber hinaus, dass die Befunde häufig widersprüchlich sind und keine Generalisierung ermöglichen. Evidenzbasierte Entscheidungen zum weiteren Umgang mit Inspektionen, zu

Veränderungs- und Verbesserungsmöglichkeiten, aber auch zur Einbettung in die länderspezifischen Bildungssysteme sind daher kaum möglich. Aktuelle Veränderungen basieren entsprechend eher auf anekdotischer Evidenz und auf Alltagserfahrungen, denn auf empirischer Evidenz.

Um die Wirksamkeit von Inspektionen jedoch nachhaltig und erfolgreich zu optimieren, sind Entscheidungsträger auf belastbare Befunde aus empirischen Untersuchungen und Evaluationen zu Wirkungen von Inspektion auf erwartete Zielkriterien wie Schul- und Unterrichtsqualität, aber auch Schülerleistungen angewiesen. Wir denken, dass der von uns vorgestellte konzeptionelle Rahmen zur Evaluation von Schulinspektionswirkungen dazu beitragen kann, derartige Informationen zu generieren, sofern er systematisch, im Sinne eines Forschungsprogramms angewendet wird.

Besonders wichtig erscheint es uns dabei, an die aktuelle Validitätsdebatte in der empirischen Bildungsforschung anzuknüpfen und das *Interpretation/Use Argument* von Kane im Rahmen einer theoriegeleiteten Evaluation von Schulinspektionswirkungen anzuwenden. Dieses Modell kann unserer Ansicht nach zur Evaluation der Wirksamkeit von Schulinspektionen genutzt werden und Programm-, Theorie- und Methodenfehler aufdecken bzw. vorbeugen. Dabei wird die Argumentationskette zweimal durchlaufen. In der ersten Phase wird die Maßnahme, in diesem Fall die Schulinspektion, evaluiert. Bei der Bewertung kann überprüft werden, ob mit den Instrumenten ermittelte Ergebnisse das Verhalten der Schule tatsächlich repräsentieren oder ob beispielsweise die Bewertung der Fragebögen, Interviews, Schulbegehungen etc. subjektiv oder fehlerbehaftet ist. Bei der Generalisierung kann untersucht werden, ob sich die Schulinspektion auf andere Facetten und Kontexte der Durchführung übertragen lässt. Hier stellt sich unter anderem die Frage, ob eine Inspektion mit anderen Inspektoren, zu einem anderen Zeitpunkt, in anderen Unterrichtsstunden oder bei anderen Lehrkräften ein gleichwertiges Ergebnis produziert hätte. Bei der Extrapolation ist es möglich, zu analysieren, ob das Konstrukt einer „guten Schule“ empirisch bestätigt werden kann und ob die Erfüllung dieser Normen tatsächlich eine „gute Schule“ in der realen Welt bedingt. Lassen sich keine Argumente oder sogar Gegenargumente dafür finden, dass das Ergebnis der Schulinspektion die reale Schulwirklichkeit beschreibt, kann dies keinen oder einen unerwünschten Prozess zur Folge haben. Gründe können ein Programm- oder ein Methodenfehler in der Schulinspektion sein. Können jedoch Argumente gefunden werden, dass das Ergebnis der Schulinspektion die reale Schulwirklichkeit beschreibt, so liegt eine solide Basis für die Schulentwicklung vor. Auf dieser Basis können dann Entscheidungen getroffen und ein gewünschter Prozess initiiert werden. Der Prozess führt wiederum zu einem eventuell veränderten beobachtbaren Verhalten der Schule. Dieses Verhalten kann erneut mithilfe von Instrumenten gemessen werden, womit das Modell von Kane ein zweites Mal durchlaufen werden kann. Bei der Bewertung kann

erneut untersucht werden, ob die mit den Instrumenten generierten Ergebnisse das Verhalten der Schule korrekt abbilden. Bei der Generalisierung ist es möglich, festzustellen, ob die Befunde aus verschiedenen Untersuchungen einheitlich und übertragbar sind. Bei der Extrapolation kann wiederum evaluiert werden, ob sich die Theorie über den Prozess durch die Testergebnisse bestätigen lässt und ob der Prozess zu erwünschten Effekten in der Wirklichkeit führt. Hierbei können Theorien und Ergebnisse aus dem Handlungsmodell und Veränderungsmodell (Chen 2006) integriert werden. Deuten die Argumente darauf, dass mit dem Instrument nicht die gewünschten Effekte des Prozesses in der Wirklichkeit repräsentiert werden, so kann es sich um einen Theoriefehler des Prozesses oder einen Methodenfehler in der Messung des Prozesses handeln. Bei der Entscheidung geht es wiederum um die Nutzung der Ergebnisse. In diesem Schritt stellt sich u. a. die Frage, ob die Maßnahme als erfolgreich gewertet werden kann, ob sie positive oder negative Konsequenzen für die Schule, die Lehrkräfte oder Schülerinnen und Schüler hat, etc. In diesem Schritt können auch die Aspekte aus dem Modell zur Kategorisierung von Inspektionswirkungen berücksichtigt werden, beispielsweise durch die Kategorisierung von beabsichtigten und unbeabsichtigten Prozessen und Ergebnissen über die Zeit.

Auf diese Weise bietet das Modell von Kane die Möglichkeit, Schulinspektion und ihre Konsequenzen umfassend zu evaluieren. Die Evaluation der einzelnen Schritte ist transparent und offenbart die evidenzbasierten Argumente, jedoch auch Argumente mit schwacher Evidenz, Gegenargumente und Ausnahmen. Auf diesem Wege wird es möglich, wissensbasierte Entscheidungen zur weiteren Ausgestaltung von Schulinspektionsverfahren zu treffen und das Verfahren so zu optimieren, dass es die beabsichtigten Wirkungen auch tatsächlich entfalten kann, also wirksam wird.

## Literatur

- van Ackeren, I. & Klemm, K. (2009). *Entstehung, Struktur und Steuerung des deutschen Schulsystems*. Wiesbaden: VS Verlag für Sozialwissenschaften.
- Alkin, M. C. & Taut, S. M. (2003). Unbundling evaluation use. *Studies in Educational Evaluation*, 29, 1–12.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (1999). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.
- American Educational Research Association, American Psychological Association, National Council on Measurement in Education (2014). *Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

- Astbury, B. & Leeuw, F.L. (2010). Unpacking black boxes. Mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31 (3), 363–381.
- Bachman, L.F. (2005). Building and Supporting a Case for Test Use. *Language Assessment Quarterly*, 2 (1), 1–34. doi:10.1207/s15434311laq0201\_1
- Böttcher, W. & Kotthoff, H.-G. (2007). Gelingensbedingungen einer qualitätsoptimierenden Schulinspektion. In W. Böttcher & H.-G. Kotthoff (Hrsg.), *Schulinspektionen: Evaluation, Rechenschaftslegung und Qualitätsentwicklung* (S. 223–230). Münster: Waxmann.
- Böttcher, W. & Kotthoff, H.-G. (2010). Neue Formen der Schulinspektion: Wirkungshoffnungen und Wirksamkeit im Spiegel empirischer Bildungsforschung. In H. Altrichter & K. Maag Merki (Hrsg.), *Neue Steuerung im Schulwesen* (S. 295–325). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Brunsdén, V., Davies, M. & Shevlin, M. (2006). Anxiety and stress in educational professionals in relation to OFSTED. *Education Today*, 56 (1), 24–31.
- Chen, H.T. (2006). A theory-driven evaluation perspective on mixed methods research. *Research in the Schools*, 13 (1), 75–83.
- Coryn, C.L.S., Noakes, L.A., Westine, C.D. & Schröter, D.C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32 (2), 199–226.
- Cronbach, L.J., Ambron, S.R., Dornbusch, S.M., Hess, R.D., Hornik Phillips, D.C., Walker, D.F. & Weiner, S.S. (1980). *Toward reform of program evaluation*. San Francisco: Jossey-Bass.
- Dederling, K. (2012). Schulinspektion als wirksamer Weg der Systemsteuerung? *Zeitschrift für Pädagogik*, 58 (1), 69–88.
- de Wolf, I.F. & Janssens, F.J.G. (2007). Effects and side effects of inspection and accountability in education: on overview of empirical results. *Oxford Review of Education*, 33 (3), 379–396.
- Döbert, H., Rürup, M. & Dederling, K. (2008). Externe Evaluation von Schulen in Deutschland – die Konzepte der Bundesländer, ihre Gemeinsamkeiten und Unterschiede. In H. Döbert & K. Dederling (Hrsg.), *Externe Evaluation von Schulen. Historische, rechtliche und vergleichende Aspekte* (S. 63–152). Münster: Waxmann.
- Ehren, M.C.M., Leeuw, F.S. & Scheerens, J. (2005). On the impact of the Dutch educational supervision act: Analysing assumptions concerning the inspection of primary schools. *American Journal of Evaluation*, 26 (1), 60–76.
- Ehren, M.C.M. & Visscher, A.J. (2006). Towards a theory on the impact of school inspections. *British Journal of Educational Studies*, 54 (1), 51–72.
- Ehren, M.C.M. & Visscher, A.J. (2008). The relationships between school inspections, school characteristics and school improvement. *British Journal of Educational Studies*, 56 (2), 205–227.
- Herbert, J.L. (2014). Researching evaluation influence: A review of the literature. *Evaluation Review*, 38 (5), 388–419.

- Husfeldt, V. (2011). Wirkungen und Wirksamkeit der externen Schulevaluation: Überblick und Stand der Forschung. *Zeitschrift für Erziehungswissenschaft*, 14 (2), 259–283.
- Kane, M. T. (2002). Validating high-stakes testing programs. *Educational Measurement: Issues and Practice*, 21 (1), 31–41.
- Kane, M. T. (2006). In Praise of Pluralism. A Comment on Borsboom. *Psychometrika*, 71 (3), 441–445. Verfügbar unter: <http://doi.org/10.1007/s11336-006-1491-2> [Zugriff am 30.03.2015]
- Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, 50 (1), 1–73. Verfügbar unter: <http://doi.org/10.1111/jedm.12000> [Zugriff am 30.03.2015]
- King, J. A. & Pechman, E. M. (1984). Pinning a wave to the shore: Conceptualizing evaluation use in school systems. *Educational Evaluation and Policy Analysis*, 6 (3), 241–451.
- Kirkhart, K. E. (2000). Reconceptualising evaluation use: An integrated theory of influence. *New Directions for Evaluation*, 88, 5–23.
- Kluger, A. N. & DeNisi, A. S. (1996). The Effects of Feedback Interventions on Performance: Historical Review, a meta-analysis and a preliminary feedback intervention theory. *Psychological Bulletin*, 119 (2), 254–284.
- Köller, O. (2009). Evaluation pädagogisch-psychologischer Maßnahmen. In E. Wild & J. Möller (Hrsg.), *Pädagogische Psychologie* (S. 331–351). Heidelberg: Springer.
- Luginbuhl, R., Webbink, D. & de Wolf, I. F. (2009). Do inspections improve primary school performance? *Educational Evaluation and Policy Analysis*, 31 (3), 221–237.
- Maritzen, N. (2009). Schulinspektion und Schulaufsicht. In H. Buchen & H.-G. Rolff (Hrsg.), *Professionswissen Schulleitung* (S. 1368–1392). Weinheim: Beltz.
- Messick, S. (1998). Test Validity: A Matter of Consequence. *Social Indicators Research*, 45 (1/3), 35–44. Verfügbar unter: <http://doi.org/10.1023/A:1006964925094> [Zugriff am 30.03.2015]
- Mislevy, R. J. (2003). Substance and structure in assessment arguments. *Law, Probability, and Risk*, 2, 237–258.
- Müller, S. (2010). Erste Effekte von Schulinspektion – eine Zwischenbilanz. In N. Berke-meyer, W. Bos, W., H. G. Holtappels, N. McElvany & R. Schulz-Zander (Hrsg.), *Jahrbuch der Schulentwicklung Band 16. Daten, Beispiele und Perspektiven* (S. 289–308). Weinheim: Juventa.
- Nutley, S., Walter, I. & Davies, H. T. O. (2003). From Knowing to doing: A framework for understanding the evidence-into-practice agenda. *Evaluation*, 9 (2), 125–148.
- Pietsch, M., Janke, N. & Mohr, I. (2013). Führt Schulinspektion wirklich nicht zu besseren Schülerleistungen? Eine Einschätzung zur Belastbarkeit vorliegender Wirksamkeitsstudien aus programmtheoretischer Perspektive. In K. Schwippert, M. Bonsen & N. Berke-meyer (Hrsg.), *Schul- und Bildungsforschung. Diskussionen, Befunde und Perspektiven* (S. 167–185). Münster: Waxmann.
- Pietsch, M., Janke, N. & Mohr, I. (2014). Führt Schulinspektion zu besseren Schülerleistungen? Difference-in-Differences-Studien zu Effekten der Schulinspektion

- Hamburg auf Lernzuwächse und Leistungstrends. *Zeitschrift für Pädagogik*, 60 (3), 446–470.
- Pietsch, M., Schnack, J. & Schulze, P. (2009). Unterricht zielgerichtet entwickeln: Die Schulinspektion Hamburg entwickelt ein Stufenmodell für die Qualität von Unterricht. *Pädagogik*, 61 (2), 38–43.
- Pietsch, M., Schulze, P., Schnack, J. & Krause, M. (2011). Elaborierte Rückmeldungen zur Qualität von Unterricht. Über empirisch abgesicherte Bezugsnormen als Grundlage für die Weiterentwicklung von Unterricht und Schule. In S. Müller, M. Pietsch & W. Bos (Hrsg.), *Schulinspektionen in Deutschland – Eine Zwischenbilanz aus empirischer Sicht* (S. 193–216). Münster: Waxmann.
- Ramaprasad, A. (1983). On the Definition of Feedback. *Behavioral Science*, 28 (1), 4–13.
- Scheerens, J. (2007). The case of evaluation and accountability provisions in education as an area for the development of policy malleable system indicators. In H.-H. Krüger, T. Rauschenbach, U. Sander (Hrsg.), *Bildungs- und Sozialberichterstattung* (Zeitschrift für Erziehungswissenschaft: Sonderheft 6, S. 207–224). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Scriven, M. (2012). *Evaluating Evaluations: A Meta-Evaluation Checklist*. Verfügbar unter: [http://michaelscriven.info/images/EVALUATING\\_EVALUATIONS\\_8.1.11.pdf](http://michaelscriven.info/images/EVALUATING_EVALUATIONS_8.1.11.pdf) [Zugriff am 30.03.2015]
- Shaw, S. & Crisp, V. (2012). An approach to validation: Developing and applying an approach for the validation of general qualifications. *Research Matters: A Cambridge Assessment Publication* (Special Issue 3). Verfügbar unter: <http://www.cambridgeassessment.org.uk/images/110003-research-matters-special-issue-3-an-approach-to-validation.pdf> [Zugriff am 30.03.2015]
- Stame, N. (2010). What doesn't work? Three failures, many answers. *Evaluation*, 16 (4), 371–387.
- Stamm, M. (2003). Evaluation im Spiegel ihrer Nutzung: Grande idée oder grande illusion des 21. Jahrhunderts? *Zeitschrift für Evaluation*, 2 (2), 183–200.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Toulmin, S. (2003). *The uses of argument* (aktualisierte Aufl.). Cambridge: Cambridge University Press.
- Visscher, A.J. & Coe, R. (2003). School Performance Feedback Systems: Conceptualisation, Analysis and Reflection. *School Effectiveness and School Improvement*, 14 (3), 321–349.
- Weber, K. (2006). From nuts and bolts to toolkits: theorizing with mechanisms. *Journal of Management Inquiry*, 15 (2), 119–123.
- Weiss, C.H. (1980). Knowledge creep and decision accretion. *Knowledge: Creation, Diffusion, Utilization*, 1 (3), 381–404.
- Weiss, C.H. (1998). Have we learned anything new about the use of evaluation? *American Journal of Evaluation*, 19 (1), 21–33.